



苏州中学园区校西安交通大学少年班

信息技术项目论文
基于 U-transformer 架构的人工智能气象预测

陈尔奇¹, 朱恩泽¹, 郝田¹
指导教师: 赵俊²
ERQI CHEN¹, ENZE ZHU¹, TIAN HAO¹
Supervisor: JUN ZHAO²
2024年6月

¹ XJTU-YGP
² SHSSIP

摘要

当今的气象预测存在着一些缺陷。气象预测的人工数值预报方法需要计算大量的偏微分方程，计算量极大，且由于气象系统对初值的敏感性，其计算需大量专业技巧，即大量专业人员。AI 气象预测的方案在较简单的架构下，因模型的复杂度有限而无法很好的拟合，难以实现中期高精度预测，而复杂的架构则有巨量的算力需求。因此，较低计算量的气象预测成了空白。本文研究了如何通过 U-transformer 架构，将 U-net 融入 transformer，在较低计算量下实现全球气象高精度预测，并取得了一定的成果。

关键词

人工智能；气象预测；U-net；Swin Transformer；U-transformer 架构

目 录

摘 要	1
一、 引言	3
二、 方法	4
2.1 数据	4
2.2 模型架构	4
2.2.1 立方嵌入	5
2.2.2 U-transformer 块	5
2.2.3 全连接层	5
2.3 模型训练	6
2.4 评估方法	6
三、 研究成果	7
四、 总结	8
附录	9
数据可用性	9
代码可用性	9
参考文献	10

一、引言

准确的高精度气象预报在人类的生产生活中发挥了重要的作用。目前，各国气象局均将气象数值预报方法 (NWP) 作为预报方法之一。传统数值预报中，欧洲气象中心 (ECMWF) 实现的全球高精度中期气象预报被认为是极精确的，能够以经纬度 $0.1^\circ \times 0.1^\circ$ 的水平分辨率在 137 个高程 (这里指压力层，下文同) 上实现 10 天的逐小时有效预报，或以 18 千米的水平分辨率在 91 个高程上实现 15 天的逐小时有效预报。与此同时，采用人工智能进行全球高精度气象预测也有了不错的发展。盘古气象大模型利用 3DEST (3D Earth Special transformer) 首次将 AI 气象预报的准确率提至人工数值预报方法之上，展现了 transformer 在这一领域的潜力。GraphCast 采用图神经网络 (GNN)，以较低的计算量实现了极佳的效果。伏羲气象大模型将 U-net 融入 transformer，以低于盘古的计算量达到了不错的准确率。以上的 AI 气象模型均是以 0.25° 的水平分辨率在几十个高程上实现高精度的全球气象预测。

然而，现今的气象预报领域仍存在一定的问题。在人工数值预报方法中，为了求解众多的偏微分方程，需要执行庞大的计算任务。同时，由于天气系统对初始条件的高度敏感性，这些计算过程需要丰富的专业技能和人力资源。在 AI 气象预报方面，虽然简单模型架构因为模型复杂度不足，难以实现中期的精确预报，而复杂模型架构则对计算资源提出了极高的要求。因此，在低计算成本气象预报方面存在一定的空白。

为探索轻量化 transformer 在气象预测领域的潜力，我们简化了伏羲气象模型的 U-transformer 架构，在其基础上进行创新，研究了如何利用 transformer 用较低计算量实现全球气象的高精度预测，并取得了一定的成果。具体来说，本文利用 U-transformer 架构，使用少于伏羲的算力，以 0.25° 的水平分辨率在 10 个高程上实现了 6 小时 5 天左右的有效预测。

二、方法

2.1 数据

我们采用的数据集来源于欧洲气象中心的第五代再分析数据集，即 ERA5 数据集。ERA5 数据集包含每小时一次的 1940 年 1 月至今在 136 个高程上的高空数据、1950 年 1 月至今的地表数据，其水平分辨率约为 31 千米。ERA5 数据集被认为是最全面、最精确的全球气象数据集，且其数据的地点选取较为规律和均匀，便于后续的处理，因此，我们将 ERA5 数据集作为模型训练和测试所依赖的实际数据。

我们取 ERA5 数据集中每天 0 时、6 时、12 时和 18 时的数据，采取的水平分辨率为 0.25° ，即选择 1440 个经度和 721 个纬度上 721×1440 个地点的数据。在每一时刻的每一地点，我们在地面数据中选择了距地 2 米处气温 (2m temperature)、地面气压 (Surface pressure)、总降水量 (Total precipitation)、距地 10 米处经向风速 (10m u-component of wind)、距地 10 米处纬向风速 (10m v-component of wind) 这 5 个数据，在高空数据中选择了 9 个高程 (100、200、300、400、500、600、700、850、1000hPa) 上的气温 (Temperature)、海拔高度 (Geopotential)、相对湿度 (Relative humidity)、经向风速 (U-component of wind)、纬向风速 (V-component of wind) 各 5 个数据，共 50 个数据 ($50=5+9 \times 5$)，作为预测的关注对象。将其合并，则可得到每个时刻的 $50 \times 721 \times 1440$ 个数据。

由于算力和数据下载时间的限制，在 ERA5 数据集中，我们仅把 2024 年 1 月、2 月、3 月的数据作为训练集，将 4 月、5 月的数据作为测试集和验证集。

2.2 模型架构

由于气象系统具有马尔科夫性，即气象系统下一时刻的状态由且仅由这一时刻的状态决定，理论上仅输入一个时刻的数据即可进行预测。但考虑到人类可得数据的不完整和误差，以及气象系统的混沌特点，我们采取了输入 2 个时刻的数据 ($2 \times 50 \times 721 \times 1440$) 进行预测的方案这一合情的选择。而要生成一段时间的预测，模型先利用 $t-1$ 时刻、 t 时刻的数据预测 $t+1$ 时刻的数据，再利用 t 时刻、 $t+1$ 时刻的数据预测 $t+2$ 时刻的数据，以此类推，即可实现中期的气象预测。

我们采用的 U-transformer 架构由立方嵌入 (Cube Embedding)、U-transformer 块 (U-transformer Block) 和一个全连接层 (FC Layer) 组成，其中的 U-transformer 块将 U-net 与 transformer 相结合，由一个 U-net 中的下采样块 (Down Block)、transformer 块、一个 U-net 中的上采样块 (Up Block) 组成。另外，由于气象系统的混沌性，AI 模型会在进行 Cube Embedding 前在输入中添加柏林噪声，其有 0.5 的缩放因子，沿每个维度 (信道、纬度和经度) 产生的噪声周期数 (时间、纬度和经度) 分别为 1、6 和 6。

2.2.1 立方嵌入

立方嵌入 (Cube Embedding) 是用 3 维卷积进行下采样的层。具体来说, 是用 C 个 $2 \times 4 \times 4$ 的卷积核, 以 $2 \times 4 \times 4$ 的步长, 将尺寸为 $2 \times 50 \times 721 \times 1440$ 的数据处理为尺寸为 $C \times 180 \times 360$ 的数据。 C 即通道数, 也是卷积核的数量, 我们选取的 C 值为 512。在此之后, 数据会被归一化再输入 U-transformer 块, 以提升模型训练的稳定性。这一层有卷积层的提取特征的作用, 提取了全球气象系统中如锋面系统的小系统的特征, 交由后续层进行数据处理, 且降低了计算量, 减轻了训练的算力负担。

2.2.2 U-transformer 块

U-transformer 块融合了 U-net 和 transformer 的架构, 由一个 U-net 中的下采样块 (Down Block)、transformer 块、一个 U-net 中的上采样块 (Up Block) 组成。尺寸为 $C \times 180 \times 360$ 的数据经过下采样块变为 $C \times 90 \times 180$, 经过 24 个重复的不改变数据尺寸的 transformer 块, 再与下采样块的输出合并 (类似 U-net), 经上采样块变回 $C \times 180 \times 360$ 。

其中, 下采样块在 U-net 的基础上进行了修改。U-net 的下采样块由两个卷积层和一个池化层组成。为了减少数据的损失, 降低计算量, 我们将 U-net 下采样块的第二个卷积层经度和纬度这两个维度的步长设置为 2, 删去了池化层, 替代了其作用。下采样块后跟随着一个组归一化层和一个 SiLU 激活函数, 以增强模型的稳定性和性能。上采样块与 U-net 相同, 由两个卷积层和一个上采样组成。

模型使用的 transformer 块是 swin transformer v2 中的 transformer 块。swin transformer 使用 Window Attention, 让每一处的输出重点关注其周围地点的输入, 符合气象系统一处的气象变化仅取决于周围气象的性质。swin transformer v2 相较于 swin transformer (v1), 将层归一化从注意力层前改为注意力层之后, 并采取含偏置的余弦注意力, 在气象预测这类复杂任务拥有更强的性能和更好的稳定性。

swin transformer 的相对位置偏差来表示注意力的常量部分, 是根据每个窗口的相对坐标计算得出的。然而, 对于全球天气预报来说, 情况有所不同。有些天气状态与绝对位置密切相关, 例如拉尼娜现象、台风和飓风。每个数据都对应于地球坐标系上的一个绝对位置。因为地图是地球球面的投影、相邻标记之间的间距可能不同。为了捕捉这些特性, 我们引入了它的工作原理是, 根据每个标记的绝对位置, 为其添加一个可学习的绝对偏置。

2.2.3 全连接层

U-transformer 块输出注意力数据后, 经由全连接层得到预测结果。全连接层即是多维的单层全连接层, 但由于在 Cube Embedding 的过程中, 数据中有效的纬度数由 721 变为了 720, 因此全连接层的输出应为 720, 与输入相统一并降低计算量。输出数据被整形为 $50 \times 720 \times 1440$ 的尺寸后, 经双线性插值, 将其改变为 $50 \times 721 \times 1440$, 即下一时刻的全球气象数据的预测结果。

2.3 模型训练

我们采用了 L1 类型的损失函数 (L1 Loss)，如下：

$$L1(t) = \frac{1}{C \times H \times W} \sum_{c=1}^C \sum_{i=1}^H \sum_{j=1}^W |\hat{X}_{c,i,j}^{t_0+t} - X_{c,i,j}^{t_0+t}|$$

其中， C ， H ， W 分别代表输出的通道数（数据类数）、维度数、经度数， $\hat{X}_{c,i,j}^{t_0+t}$ 代表 t 个时间步长后的预测数据， $X_{c,i,j}^{t_0+t}$ 代表 t 个时间步长后实际数据。

目标为全球气象预测的 U-transformer 模型的训练分为两个部分：预训练和分步进行微调。训练均采用 AdamW 优化器，参数 $\beta_1=0.9$ ， $\beta_2=0.95$ ，权重衰减为 0.05，采用 16 位精度训练，批大小 (batch size) 为 1。

预训练采取一个时间步长（6 小时）的损失函数，使用的学习率为 0.001，训练 60 轮 (epoch)，使得模型在一个时间步长下的预测达到一定水平，这是较长时间的全球气象预测的基础。

在结束预训练后，对基础模型进行第一步微调，将损失函数调整至 1-10 个时间步长上 L1 Loss 的平均值，学习率为 0.0001，训练 25 轮，以获得 0 到 2.5 天 的 6 小时预测的最佳性能。接着对模型进行第二步微调，将损失函数调整至 1-20 个时间步长上 L1 Loss 的平均值，学习率为 0.0001，训练 15 轮，以获得 0 到 5 天 的 6 小时预测的最佳性能。这样分两步进行微调，既能有效提升模型中期预测的性能，又因为一定性能下的模型的训练收敛更快、计算量更小，降低了模型训练对算力的要求。

2.4 评估方法

我们用如下两个函数评估模型预测的精确度：

$$ACC(c, t) = \frac{\sum_{i,j} (\hat{X}_{c,i,j}^{t_0+t} - M_{c,i,j}^{t_0+t})(X_{c,i,j}^{t_0+t} - M_{c,i,j}^{t_0+t})}{\sqrt{\sum_{i,j} (\hat{X}_{c,i,j}^{t_0+t} - M_{c,i,j}^{t_0+t})^2} \sqrt{\sum_{i,j} (X_{c,i,j}^{t_0+t} - M_{c,i,j}^{t_0+t})^2}}$$

$$RMSE(c, t) = \frac{1}{H \times W} \sum_{i,j} (\hat{X}_{c,i,j}^{t_0+t} - X_{c,i,j}^{t_0+t})^2$$

其中， M 表示实际数据的均值。

准确率函数 (ACC) 代表了预测数据与实际数据的正相关程度，当预测数据与平均数据只差和实际数据与平均数据之差成正比时，准确率达 1。均方根误差 (RMSE) 代表了预测数据与实际数据的偏差大小。

三、研究成果

在一张 4060GPU 上训练了 3 天后，模型完成了训练。我们将模型在测试集上的 Z500（500hPa 处的海拔高度）、T2M（距地 2 米处的气温）、T850（850hPa 处的气温）以及 U10（距地 10 米处的经向风速）的均方根误差和准确率作为评价指标。我们取 6 小时、72 小时、120 小时的数据

根据数据，我们可以得出以下结论：

1. 均方根误差 (RMSE) 分析：

对于 Z500，RMSE 值从 6 小时的 157.1 增加到 72 小时的 894.4，再到 120 小时的 1388.3。这表明预测的不确定性随时间增加而显著增加。T2m 的 RMSE 从 6 小时的 2.21 增加到 72 小时的 4.83，再到 120 小时的 7.19，显示出预测误差的逐渐增大。T850 和 U10 的 RMSE 也表现出类似的增长趋势，分别从 1.33 和 1.56 增加到 5.73 左右。

2. 准确率 (ACC) 分析：

Z500 的 ACC 从 6 小时预报的 0.99 下降到 120 小时预报的 0.91，表明预测的准确性随时间增加而显著下降。T2m 的 ACC 从 6 小时的 0.99 下降到 120 小时的 0.92，显示出预测准确性的逐渐降低。T850 和 U10 的 ACC 也显示出随时间增加而下降的趋势，T850 的 ACC 由 0.96 下降至 0.87，U10 的准确率下降最为显著，从 6 小时的 0.95 降至 120 小时的 0.42。

这些数据反映了天气预报模型在不同时间尺度上的性能。在较短的时间尺度（6 小时）上，模型表现出较高的准确性和较小的预测误差。然而，随着预报时间的延长（72 小时和 120 小时），模型的准确性和预测精度显著下降，预测误差显著增大。这是由于大气动力学的复杂性以及初始条件和边界条件的误差累积。U10 等数据的误差可能较大，对于中长期天气预报会有较大的影响。

总而言之，U-transformer 在全球中期高精度气象预报中可以实现 5 天左右的有效预报，这主要是因为一些数据的预测的短板。

四、总结

目前，人工智能气象预测在轻量化高精度中期预测较为空白，仅 GraphCast 做出了一定的探索。我们在伏羲气象大模型提出的 U-transformer 架构的基础上，作出了一定的创新和改进，并实现 5 天左右的有效预报，研究了 transformer 在这一领域的潜力。

我们有一些创新之处。我们修改了 U-transformer 块中下采样块的第二个卷积层和池化层，将其合并，以降低计算量、提升模型性能。我们将 swin transformer 块的相对偏置修改为绝对偏置，以适应地球现实，并降低计算量。我们在伏羲的训练方法上进行了修改，将三个联级模型更改为在一个模型上分两步完成训练。

我们的研究还有一些不足：一，计算量依然有降低的空间，这有待进一步的研究；二，我们的预测结果显现出少数短板制约模型预测精确度的现象，修改损失函数可能可以解决这一问题；三，模型在较长的时间步长下表现不佳，可能需要另外的训练技巧，或者直接把时间步长改为 48 小时，结合短期预测模型实现较高精度的预测；四，我们的模型依然不是完全端到端的，需要再分析数据的支持。

在完成项目的过程中，我们学习了部分经典模型架构，阅读了一定量的源码和文献，了解了一些机器学习和信息处理的库和开源项目，提升了自己在这一领域的能力和小组合作完成项目的的能力。本次项目实践增强了我们对机器学习和人工智能的认识，同时也让我们更加全面的理解了机器学习的基本原理和设计方法和参数的作用。

附录

数据可用性

我们从哥白尼气候数据(CDS)官方网站下载了 ERA5 数据集子集,在气候数据(CDS)的官方网站 <https://cds.climate.copernicus.eu/>上下载了 ERA5 数据集子集。

代码可用性

我们使用和修改了 swin transformer v2 的源码,形成了 U-transformer 块中的 transformer 块,请访问 <https://github.com/microsoft/Swin-Transformer>。柏林噪声的实现基于 GitHub 上的开源代码: <https://github.com/pvigier/perlin-numpy>。

参考文献

- [1]. Haiden, T. et al. Evaluation of ECMWF forecasts, including the 2021 upgrade (2021).
- [2]. Bi, K. et al. Accurate medium-range global weather forecasting with 3d neural networks. *Nature* (2023).
- [3]. Lam, R. et al. Graphcast: Learning skillful medium-range global weather forecasting (2022). <https://arxiv.org/abs/2212.12794>.
- [4]. Chen, L., Zhong, X., Zhang, F. et al. FuXi: a cascade machine learning forecasting system for 15-day global weather forecast. *npj Clim Atmos Sci* **6**, 190 (2023). <https://doi.org/10.1038/s41612-023-00512-1>.